

Propuesta de Solución de Automatización:

Herramientas o Scripts para Extraer Datos de las Hojas de Cálculo:

Google Sheets API: Utilizar la API de Google Sheets para acceder y extraer datos de las hojas de cálculo de Google. Esto permitirá la automatización de la obtención de datos directamente desde las hojas de cálculo en la nube.

Python Script: Desarrollar un script en Python para interactuar con la API de Google Sheets. Se pueden utilizar bibliotecas como `gspread` para facilitar la extracción de datos desde las hojas de cálculo.

Proceso de Mapeo de Datos y Transformaciones:

Definir Mapeo de Campos: Identificar y documentar cómo se mapearán los campos de las hojas de cálculo al esquema de la base de datos. Es importante tener un mapeo claro para garantizar la consistencia de los datos.

Aplicar Transformaciones: Implementar transformaciones necesarias para adaptar los datos al formato esperado por la base de datos. Esto puede incluir cambios de formato, limpieza de datos y conversión de tipos.

Manejo de Actualizaciones y Duplicados:

Identificación de Duplicados: Antes de la carga, realizar una comparación con los datos existentes en la base de datos para identificar posibles duplicados. Esto se puede hacer mediante la comparación de identificadores únicos o combinaciones de campos.

Manejo de Actualizaciones: Decidir cómo manejar actualizaciones en los eventos existentes. Puede ser necesario definir reglas sobre qué datos se deben actualizar y en qué condiciones.

Programación de Importaciones, Frecuencia, Monitoreo de Errores:

Programación de Importaciones: Establecer un cronograma para la ejecución automática del script de importación. La frecuencia dependerá de la frecuencia de actualización de las hojas de cálculo.

Registro y Monitoreo de Errores: Implementar un mecanismo de registro para rastrear cualquier error durante el proceso de importación. Esto facilitará la identificación y resolución de problemas.

Documentación:

Modelo de Datos y Script de Carga:

Incluir el modelo de datos SQL diseñado en el repositorio.
Proporcionar un script de carga inicial que refleje la lógica de mapeo y transformación de datos.

Documento de Análisis y Diseño para la Automatización:

Describir los requisitos del sistema, incluyendo la frecuencia de actualización de los datos.
Detallar el diseño de la solución propuesta, explicando la lógica de mapeo, transformaciones y manejo de duplicados y actualizaciones.
Proporcionar una guía clara sobre cómo se utilizarán las herramientas y scripts propuestos.

Esta propuesta se centra en la utilización de herramientas estándar y un enfoque programático para garantizar una carga de datos automatizada, consistente y eficiente.

Proceso de Mapeo de Datos y Transformaciones:

Definir Mapeo de Campos:

Identificar Campos en Hojas de Cálculo:

- Revisa las hojas de cálculo disponibles y documenta todos los campos presentes en cada una. Estos campos pueden incluir nombres de eventos, fechas, ubicaciones, etc.

Analizar Estructura de la Base de Datos:

- Examina la estructura de la base de datos y documenta los campos existentes en las tablas correspondientes. Asegúrate de que haya una correspondencia clara entre los campos de las hojas de cálculo y los de la base de datos.

Asignar Tipos de Datos:

- Define los tipos de datos para cada campo en la base de datos. Por ejemplo, si en la hoja de cálculo tienes una columna de fechas, asegúrate de que haya un campo en la base de datos con el tipo de dato adecuado para almacenar fechas.

Establecer Claves y Relaciones:

- Si la base de datos utiliza claves primarias o tiene relaciones entre tablas, asegúrate de entender cómo se asignarán los identificadores únicos y cómo se relacionarán las tablas.

Documentación del Mapeo:

- Crear una tabla o documento que muestre la correspondencia entre los campos de las hojas de cálculo y los de la base de datos. Esto servirá como referencia clara para la fase de implementación.

Aplicar Transformaciones:

Limpieza de Datos:

- Identificar posibles problemas de calidad de datos, como valores nulos o inconsistencias, y decidir cómo manejarlos. Se debe realizar limpieza de datos para garantizar consistencia.

Cambio de Formato:

- Si los datos en las hojas de cálculo no están en el formato esperado por la base de datos, implementar transformaciones para cambiar el formato.

Conversión de Tipos:

- Asegurarse de que los tipos de datos coincidan entre las hojas de cálculo y la base de datos. Si es necesario, realizar conversiones de tipos. Por ejemplo, si una columna está representada como texto pero debe ser numérica, realiza la conversión.

Normalización de Datos:

- Si hay redundancias en los datos que deben evitarse, considera aplicar técnicas de normalización. Esto puede incluir la creación de tablas adicionales para evitar la repetición de información.

Validación de Datos:

- Implementar reglas de validación para garantizar que los datos cumplen con ciertos criterios antes de ser cargados en la base de datos. Esto ayuda a mantener la integridad de los datos.

Mapeo de Valores:

- Si hay valores específicos en las hojas de cálculo que deben mapearse a valores específicos en la base de datos (por ejemplo, categorías o códigos), realiza estas asignaciones.

Pruebas de Transformación:

- Realizar pruebas exhaustivas de las transformaciones para asegurarte de que los datos se ajustan correctamente al esquema de la base de datos. Esto puede incluir la carga de un conjunto de datos de prueba y la verificación de los resultados.